Applications of Computational Methods in Natural Products Based Drug Discovery

2

Ahmad Abu Turab Naqvi, Syed Afzal Murtaza Rizvi, and Md. Imtaiyaz Hassan

Abstract

Since antiquity, natural resources, mainly plants, have been used for medicinal purposes. The primitive usage was based on a trial-and-error strategy. When time passed, humans started to look deeper into the actual elements responsible for the cure. They started using plant extracts as medicines. In 1806, there opened a new window in the area of natural product-based medicines when Friedrich Sertürner, a German pharmacist, isolated morphine from the poppy plants. This is just the beginning of a new era. Soon, the extraction of phytochemicals i.e., plant-based chemical compounds, became common and chemists started looking for new ways of manipulating them and synthesizing them in laboratories. Charles Frédéric Gerhardt first synthesized acetylsalicylic acid, the wonder drug Aspirin by treating acetyl chloride and sodium salicylate in 1853. This was a revolution in drug discovery which is still running towards more advancements and developments. Later on, computers came into the picture, and chemists and molecular biologists started to visualize and analyze chemical compounds. The emergence of advancements in computers set the foundation stone for a new field of science called bioinformatics. Soon began the applications of computers in medicinal research and a new trend of computer-aided drug discovery started. Which broadly changed the face of the natural product (NP) based drug discovery process. We have huge libraries of NP-based compounds utilized to discover new drugs against several life-threatening diseases, including cancer. This chapter

A. A. T. Naqvi · S. A. M. Rizvi
Department of Computer Science, Jamia Millia Islamia, New Delhi, India

M. I. Hassan (✉)
Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
e-mail: mihassan@jmi.ac.in

talks about applications of computational methods mainly belonging to Bioinformatics and Chemoinformatics that are applied towards NP-based drug discovery.

Keywords

Plants · Natural products · Drug discovery · Bioinformatics · Chemoinformatics · Machine learning

2.1   Background

Since ancient times humans have been exploring natural sources like plants and herbs for medicinal purposes. The primitive usage of these products was limited to the direct application of plant leaves, barks, extracts, fruits, seeds, etc., as medicines. The earliest traces of the medicinal usage of plants are found in the Sumerian civilization's Clay tablet records (Petrovska 2012). Historical records from ancient Egypt, Greek, and the ancient Indus valley civilization have shed light on the pre-historic usage of plants for therapeutic aspects (Bernardini et al. 2018; Dias et al. 2012). However, for a long time and still, to some extent, these techniques have been considered a clear example of medicinal witchcraft despite their success in treating certain ailments of humans and animals. The emergence of molecular biology has made some remarkable discoveries that emphasized the importance of natural resources as medicinal agents.

When we talk about natural products (NPs) in the context of drug discovery, the majority of instances come from the plant kingdom or medicinal plants. Phytochemicals are prominent examples of NPs in drug discovery. Besides, fungi, bacteria, certain insects, and other vertebrates have also been reached for medicinal purposes. There are so many comprehensive reviews that shed light in detail on the types of NPs (Cragg et al. 1997; Harvey et al. 2015; Romano and Tatonetti 2019). In the chapter, our focus will mainly be confined to plant-based NPs or phytochemicals and various techniques in Bioinformatics and Cheminformatics that have been applied so far to develop and design new drugs.

2.2   Historical Overview of Drug Discovery from Natural Products

Although the usage of natural products for the therapy of various ailments has pre-historic traces, from the modern perspective, drug discovery from NPs ways back to the nineteenth century when extraction and chemical analysis came into existence, for a comprehensive review of the usage of NPs for medicinal purposes, readers are suggested to refer to (Petrovska 2012). This journey began with the isolation of alkaloids from plants. Friedrich Sertürner, a German pharmacist, first isolated morphine from the poppy plants in 1806 (Pathan and Williams 2012). This is the first-ever instance of extraction of NP from a plant in modern pharmacology.

After a decade, some more alkaloids were extracted from the plants like ipecacuanha and strychnos in 1817 (Petrovska 2012). The credit from the commercial extraction of alkaloids, mainly morphine, goes to Heinrich Emanuel Merck in 1826 in Germany (Atanasov et al. 2015).

The second phase of advancements in the exploitation of NPs in drug discovery began with synthesizing various phytochemicals and their derived products in the laboratory. The most remarkable example of such experiments is acetylsalicylic acid, aka Aspirin, the wonder drug used as a pain killer. Charles Frédéric Gerhardt first synthesized acetylsalicylic acid by treating acetyl chloride and sodium salicylate in 1853 (ref). Sodium salicylate is a sodium salt of salicylic acid. Salicylic acid has been mainly extracted from the Willow tree and other salicin-rich plants. Willow bark is known for medicinal purposes in the ancient Sumerian and Egyptian civilizations (Norn et al. 2009). Later on, the commercial synthesis of Aspirin began in 1897 by Felix Hoffmann in the Bayer Company.

Moreover, the discovery and synthesis of Aspirin from plant-based extracts have ignited a rapid growth and development of NP-based drug discovery, which is still on the path of progress with modern strategies and techniques. In the next section, we will discuss some strategies for NP-based drug discovery, mainly focusing on the computational and data-driven methods using Bioinformatics and Cheminformatics'.

2.3   Strategies for Drug Discovery Through Natural Products

In this section, we will discuss some modern strategies applied in NP-based drug discovery and design. These strategies make use of established Bioinformatics and Cheminformatics methods. Some leading methods incorporated in these approaches are QSAR analysis, pharmacophore modeling, molecular docking, molecular dynamics simulation, gene expression-based drug discovery, NP library construction, biomarker identification for NP-based drugs, NP-derived database development, big-data-driven drug discovery, machine learning-based methods, combinatorial library construction, fragment-based drug discovery, pharmacokinetic properties prediction and so on (see Fig. 2.1 for a generalized in silico drug discovery process).

2.3.1   Quantitative Structure-Activity Relationship (QSAR)

QSAR is a well-known technique applied in structure-based drug discovery. Although, the formal definition and technical details of QSAR are disseminated in a wide range of literature published previously. But for the brief overview for the reader, QSAR is applied to correlate the activity of a chemical compound with its structure. This uses a set of "predictor variables," which are molecular descriptors as physicochemical properties of the compounds in the context of chemical compounds. Based on these values, the biological activities of new chemical
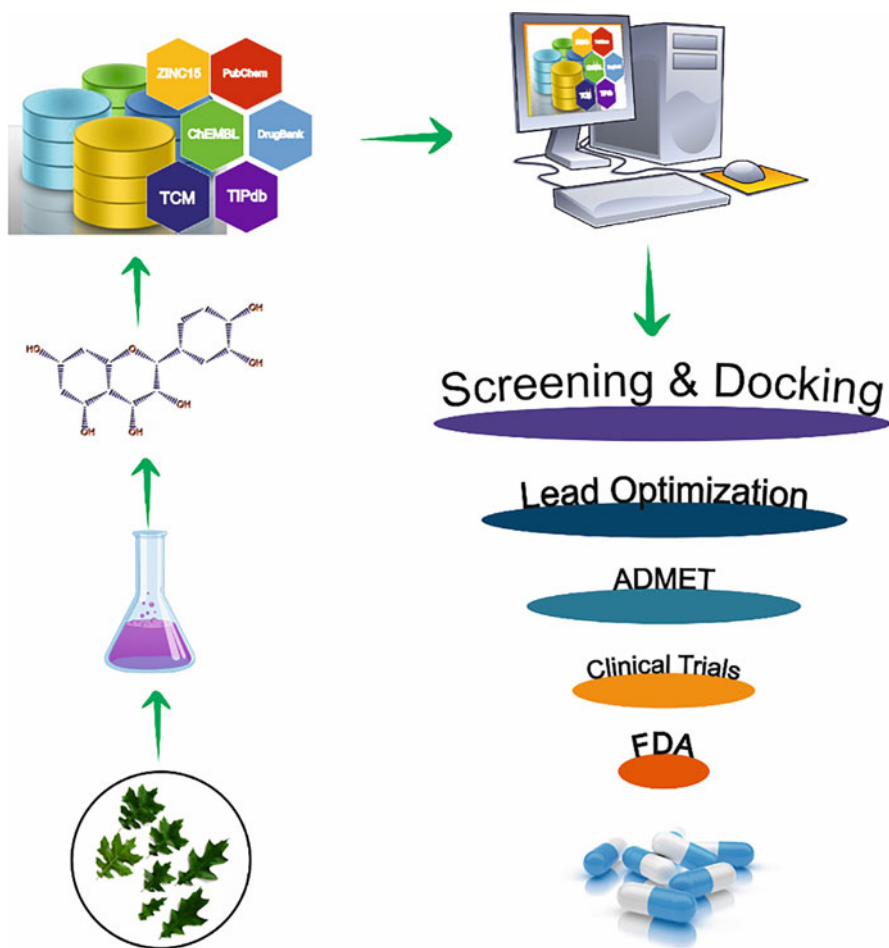
Fig. 2.1 Graphical representation of in silico drug discovery process from NPs

compounds can be predicted (Verma et al. 2010). The applications of QSAR in NP-based drug discovery we established and have come forward with promising results (Ref).

2.3.2 Pharmacophore Modelling

Pharmacophore "*is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response (IUPAC).*" Pharmacophore modeling and its application in drug discovery have seen a rise during recent years. Similarly, NP-based drug discovery has also been benefited

from the powers of pharmacophore modeling. The detailed description of pharmacophore approaches is beyond the scope of this chapter. The reader is suggested to refer to (Leach et al. 2010; Yang 2010) for a thorough understanding of the various aspects of pharmacophore modeling and its application in drug discovery.

2.3.3   High Throughput Virtual Screening and Molecular Docking

Screening of 1000 compounds against a target molecule in the laboratory is costing too much of money and time. To overcome the complications of this approach and make the processing time and cost-effective, high throughput virtual screening (HTvS) has been applied to scan a library of bioactive compounds with a biomarker for drug discovery. Subsequently, a selected set of possible candidate compounds are taken for docking analysis which further elucidates the patterns of binding and interaction with the target molecule. Docking helps identify the possible physical interaction between ligand and receptor while suggesting conformational poses that might fit into the binding cavity of the receptor. The combination of HTvS and molecular docking are well-established methods applied in computer-aided drug discovery in general and for NP-based drug development in particular.

2.3.4   Molecular Dynamics Simulations

The first successful application of MD simulation to study protein folding in 1979 opened a window for theoretical chemists to develop further the technique to study more intricate biological processes (i.e., protein–ligand interaction) at the atomic level that are otherwise difficult to study in lab conditions. With this, MD simulations became the most desired tool to study the conformational effects of ligand binding on the target molecule. MD simulations, QSAR, pharmacophore modeling, and molecular docking are widely applied in silico drug discovery studies. The mechanistic insights into the binding of a drug molecule to its target are crucial to understanding the intricate mechanism of how the drug works on its target, thus leading to the development of more specific ligands. NP-based drug discovery and development have largely benefitted from the strengths of MD simulations.

2.3.5   Prediction of Pharmacokinetic Properties and Toxicity

Knowledge of pharmacokinetic properties i.e., Adsorption, Distribution, Metabolism, and Excretion (ADME), and the safety in terms of toxicity of the NP-based compounds, is a challenge in the drug discovery process. There are several statistical and knowledge-based methods for the in silico prediction of these properties, which help eliminate the signs of "unlikely" candidates from the pool of NPs. By convention, ADME properties of a candidate molecule depend on the physicochemical

properties such as molecular weight, lipophilicity, number of hydrogen bond donors, and acceptors of the compounds. In practice, "Lipinski's Rule of Five" is applied as a filter to separate candidates that violate these parameters. Although, the application of the rule of five needs a significant amount of caution and decision making when relatively large NP-based compounds are in the picture. Hence, the rule of five has its limitations.

2.3.6   Computational Combinatorial Chemistry and Library Design

The emergence of combinatorial chemistry in the 1980s (Liu et al. 2017) opened channels for developing chemical libraries of structurally diverse chemical compounds using the structural and chemical properties of known bioactive compounds. Since then, with the advancements of computers, several methods have been developed that use combinatorial chemistry approaches to generate thousands of synthetic compounds and develop ready-to-screen chemical libraries for drug discovery purposes. In the context of NPs, this approach has been applied in recent years to develop chemical libraries of candidate compounds using naturally occurring compounds scaffolds (Grabowski et al. 2008; Mang et al. 2006). As a part of this, *computational mutagenesis* is a recent technique that involves mutating specific structural features of the target NPs to generate libraries of novel candidate compounds (Chen et al. 2002; Romano and Tatonetti 2019).

2.3.7   Machine Learning Approaches for NP-Based Drug Discovery

Machine learning algorithms have found their way through almost every domain of human life where human–machine interaction is possible. The area of computer-assisted drug discovery is also not remained untouched with applications of machine learning methods. In today's time, various aspects of drug discovery take help from the strengths of machine learning algorithms, whether in target identification or validation, de novo inhibitor design, virtual screening, docking, and ADMET property prediction (Vamathevan et al. 2019). Similar trends are also observed in NP-based drug discovery during recent years, where machine learning algorithms are applied to reduce the intricate decision-making process in drug discovery. There are discussions about possible applications of machine learning methods in predicting the functions of NPs using their two-dimensional structures (Liu et al. 2019). Moreover, manifold developments in the strengths of machine learning methods while addressing their limitations in dealing with NP-based drug discovery suggest potential success shortly.

2.3.8 Big Data and Data-Driven Drug Discovery

This section, however, is an extension to the last section, where we have discussed the applications of machine learning in drug discovery. Since these algorithms use an ample amount of predictive data that need both accuracy and integrity to reduce the possibility of false-positive prediction. In the context of drug discovery, big data refers to the huge amount of chemical information piled up in publicly accessible databases such as ZINC, PubChem, ChEMBL, and DrugBank, etc., which store millions of active compounds both naturally occurring and synthesized (Thomford et al. 2018). Big data also encircle a large amount of clinical data stored in Electronic Health Records (EHRs). Besides, disease biomarker databases, disease pathways, protein–protein interaction networks, protein–drug interaction networks, cancer gene expression data, etc., add more to the paradigm of big data. This leads to a novel but challenging aspect of drug discovery, i.e., data-driven drug discovery.

2.4 Tools and Databases for NP-Based Drug Discovery

This section will catalog various tools and databases based on or otherwise strategies above applied in NP-based drug discovery. Herein, the tools and databases discussed are generally used in computer-aided drug discovery and those sources specific to NP-based drug development. Table 2.1 provides a detailed list of available sources along with their brief description. In the following table, we have provided information on the tools and databases widely applied in the drug discovery process in general and particularly for NPs. However, the reader is suggested to refer to (Chen et al. 2017; Lagunin et al. 2014; Ma et al. 2011; Naqvi and Hassan 2017; Naqvi et al. 2018; Nguyen-Vo et al. 2020) for comprehensive details on the sources for in silico drug discovery.

2.5 Recent Advances in Drug Discovery from Natural Products

This section will explore and discuss recent case studies that have focused on natural product-based drug discovery using *state-of-the-art* in silico methods that we have discussed in the previous sections. This will direct the reader to understand and observe the current state of NP-based drug discovery. The studies discussed in this section are both purely computational technique-based and have a hybrid approach of integrating laboratory techniques with the in silico methods.

1. Monoamine oxidase B (MAO B) is associated with the catalysis of aryalkylamines neurotransmitters. Its malfunction is said to have possible involvement in the development of Parkinson's disease. In an attempt to develop potential inhibitors of MAO B, Mladenović et al. (2017) have applied, in a hybrid in vitro in silico approach, 3D-QSAR models for the evaluation of the biological activity of coumarin based compounds. Coumarin is a phytochemical which is

Table 2.1 List of tools and databased for natural product-based drug discovery

S. no.	Tools/ databases	Description	URLs
1.	AutoDock 4.2.6	It is an open-source docking tool with an improved free energy scoring function	http://autodock.scripps.edu/downloads/autodock-registration/autodock-4-2-download-page/
2.	AutoDock Vina	It is an advanced and improved version of AutoDock4	http://vina.scripps.edu/download.html
3.	FlexX	It is a highly benchmarked virtual screening and docking program for flexible docking	https://www.biosolveit.de/FlexX/
4.	Glide	Glide is a high-speed virtual screening and docking platform developed by Schrödinger	https://www.schrodinger.com/glide
5.	GOLD	It is a commercially available docking software for virtual screening, docking, and lead optimization	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
6.	LigandScout	It is suitable for pharmacophore-based screening	http://www.inteligand.com/ligandscout/
7.	SwissADME	It is a web server for the prediction of ADMET parameters. The Swiss Institute of Bioinformatics moderates it	http://www.swissadme.ch/
8.	PreADMET	A web server for ADMET property prediction	https://preadmet.bmdrc.kr/
9.	CarcinoPred-EL	It is an ensemble machine learning-based web application for carcinogenicity prediction	http://112.126.70.33/toxicity/CarcinoPred-EL/
10.	PASS online	It is a web-based biological activity prediction server. It is capable of predicting 4000 kinds of biological activities	http://www.pharmaexpert.ru/passonline/
11.	NPASS	It is a web server cum database specifically for natural product activity prediction	http://bidd.group/NPASS/
12.	Discovery studio	Commercial package for QSAR and pharmacophore modeling along with docking and molecular dynamics simulations	https://www.discngine.com/discovery-studio
13.	GUSAR	QSAR modeling and molecular descriptor prediction tool	https://genexplain.com/gusar/
14.	Molinspiration	Chemoinformatics package for various purposes, including QSAR modeling	https://www.molinspiration.com/
15.	QSARPro	Tool for 2D/3D QSAR modeling	https://www.vlifesciences.com/products/QSARPro/Product_QSARpro.php

Table 2.1 (continued)

S. no.	Tools/ databases	Description	URLs
16.	ZINC15	It is a database of commercially available compounds. It also contains a large number of natural products	https://zinc15.docking.org/
17.	PubChem	It is a database of chemical compounds and chemical activities. It contains around 111 million compounds	https://pubchem.ncbi.nlm.nih.gov/
18.	ChEMBL	It is a manually curated database of drug-like compounds. It contains around two million compounds	https://www.ebi.ac.uk/chembl/
19.	DrugBank	It is a comprehensive database of approved drugs	https://go.drugbank.com/
20.	Super Natural II	It is an exclusive database of natural products. It contains 325,508 natural compounds	http://bioinf-applied.charite.de/supernatural_new/index.php
21.	TCM Database @Taiwan	It is a database of Chinese medicines derived from natural products	http://tcm.cmu.edu.tw/
22.	CMAUP	Collective molecular activities of useful plants is a collection of around 47,000 natural compounds	http://bidd.group/CMAUP/
23.	DNP	Dictionary of natural products is a web source for commercially available natural products	http://dnp.chemnetbase.com
24.	IBS	InterBioScreen natural products library	https://www.ibscreen.com/
25.	MMPD	Myanmar medicinal plant database	https://www.tuninst.net/MMPD/MMPD-indx.htm
26.	TCMID	It is an integrated data source on traditional Chinese medicines	http://119.3.41.228:8000/tcmid/
27.	TIPdb	It is a comprehensive source of anti-cancer, anti-platelet, and anti-tuberculosis phytochemicals from indigenous plants in Taiwan	https://cwtung.kmu.edu.tw/tipdb/
28.	TMDB	It is a database of tea metabolites	http://pcsb.ahau.edu.cn:8080/TCDB/f

found in tonka beans in high concentrations. In this approach, they developed a combination of structure-based and ligand-based 3D-QSAR models and eventually deduce six relatively active inhibitors of MAO B, which might act as potential lead candidates for drug development for Parkinson's. Dhiman et al. (2018) have reviewed the application of 3D-QSAR on a diverse set of NP-based compounds such as coumarins, morpholine, piperine, naphthoquinone,

amphetamine moreover flavonoids, caffeine, and curcumin, etc., as potential MAO inhibitors. They conclude the effectiveness of QSAR and molecular docking and COMFA in finding selective and highly active inhibitors of MAO.

2. As discussed in the previous sections, molecular docking in combination with molecular dynamics simulations has proven very effective in computer-aided drug discovery studies. In recent years, many successful experiments have been conducted. These tools have been applied to discover potent and effective inhibitors for several known biomarkers of life-threatening diseases. Khan et al. (2009) studied the inhibitory effects of flavonoid derivatives quercetin, rutin, kaempferol 3-*O*-beta-D-galactoside, and macluraxanthone using molecular docking and enzyme inhibition assays against the activity of acetylcholinesterase (AChE) and butyrylcholinesterase (BChE). In this study, they found that macluraxanthone shows effective binding against both enzymes. Moreover, quercetin also exhibited strong intermolecular interactions with both these enzymes.

Cozza et al. (2006) identified ellagic acid as an effective inhibitor of casein kinase 2 (CK2) using virtual screening and molecular docking methods. In a recent study, Zhang et al. (2020) screened a library of 2080 NPs to discover their efficacy as potential antiviral compounds. The study aimed to further utilize the compounds as a potent HIV-1 capsid (CA) protein inhibitor. Based on molecular docking, they found compounds rubranol and hirsutanonol showing strong inter-molecular binding with HIV-1 CA.

Ebola virus nucleoprotein (EBOV NP) is significant for its proliferation. To develop effective inhibitors against EBOV NP, Nasution et al. (2018) screened a library of 190,084 NPs from the ZINC database. To evaluate top-scoring compounds' binding affinity and effectiveness, they applied a flexible docking approach and molecular dynamics simulation. Eventually, α-lipomycin and 3-(((*S*)-1-amino-1,2,3,4-tetrahydroisoquinolin-5-yl)methyl)-5-((5-((5*R*,7*S*)-5,7-dihydroxy-3-oxodecyl)-2-hydroxyphenoxy)methyl)pyrrolo[3,4-b]pyrrol-5-ium were found showing strong binding thus posing as potent candidates as anti EBOLA drug.

3. Machine learning algorithms coupled with QSAR or molecular docking have proven very effective in elucidating the inhibitory effects of NPs towards the identification of novel drug candidates (Korotcov et al. 2017; Lavecchia 2015). Classical MD simulation when integrated with machine learning-based methods, enhances the performance of in silico drug discovery processes (Perez et al. 2018). Shi et al. (2020) applied machine learning models to discover New Delhi metal beta-lactamase (NDM-1) inhibitors. NDM-1 producing bacteria are crucial in drug-resistant bacterial infections. They screened a library of NP-based compounds using prediction models and also compared their performance with the virtual screening and docking strategy. As a result, machine learning models exhibited 90.5% accuracy in predicting the potent inhibitors in comparison to 69.14% accuracy by the traditional docking approach.

Besides inhibitor discovery, machine learning is also applied to ADME property prediction and toxicity profiling of the NPs. In an attempt to assess the efficacy of

machine learning in the toxicity profiling of natural compounds, Onguene et al. (2018) carried out a toxicity assessment of three compound libraries of African flora, which have anti-malarial and anti-HIV activity. When compared to available experimental data for toxicity, machine-learning models were found to agree with the compounds' predicted toxicity.

4. Biological and chemical data during recent years has seen a manifold increase during recent years. Moreover, clinical data for several patients in the for EHRs is also available worldwide. This has opened a new window in the realms of drug discovery, called "data-driven drug discovery." The astronomical amount of data, preferably referred to as "big data," has a tremendous scope towards novel drug discovery (Lusher et al. 2014). During recent years, efforts have been made to study the drug effects and their interactions with pathogenic drug targets using the available information stored as clinical records and EHRs (Tatonetti et al. 2012; Yao et al. 2011). Despite the effectiveness of this approach in revolutionizing the drug discovery era, there are certain limitations, such as hurdles in accessing the clinical records or EHRs or limitations in the understanding of clinical data by informatics researchers. However, projects like Electronic Medical Records and Genomics (eMERGE) network (McCarty et al. 2011) and Observational Health Data Sciences and Informatics (OHDSI) (Hripcsak et al. 2015) are moving towards removing these barriers.

2.6   Challenges and Prospects

In this chapter we have discussed methods and strategies to find out potential drug candidates from NP-based compounds. Most of these approaches seem promising in providing relevant answers to drug discovery problems both in traditional laboratory-based methods and computational techniques. But then the question arises. Do the drug candidates and so-called "potential" inhibitors reach their destination? Destination as in as marketed drugs in the real world treating real diseases. If we see the statistics, the results are somehow satisfying. According to the survey of Newman and Cragg (2016), out of 1562 drugs approved between 1981 and 2014, 646 drugs are either NPs or NP-derived. Another survey suggests that around one-third of the new molecular entities (NMEs) approved by the FDA belong to NPs (Patridge et al. 2016). The statistics for the success of drug candidates discovered through in silico methods are also promising (Zhu et al. 2018).

Despite all the success of in silico drug discovery studies in general or in particular for NP-based compounds, some challenges still need to be addressed for a better future of NP-based drug discovery. These challenges cover the under or overuse of computational methods for discovery, unequal distribution of chemical data over several sources, thus limiting the access in most cases, controlled or no access to clinical data such as in the case of EHRs, etc. However, with the rapid development of computer hardware to mimic more intricate biological and molecular processes, processing large chemical libraries, and providing better solutions to

the challenging problems faced in computer-aided drug discovery, the future of in silico methods for NP-based drug discovery is bright and promising.

References

Atanasov AG, Waltenberger B, Pferschy-Wenzig EM, Linder T, Wawrosch C, Uhrin P et al (2015) Discovery and resupply of pharmacologically active plant-derived natural products: a review. Biotechnol Adv 33(8):1582–1614. https://doi.org/10.1016/j.biotechadv.2015.08.001

Bernardini S, Tiezzi A, Laghezza Masci V, Ovidi E (2018) Natural products for human health: an historical overview of the drug discovery approaches. Nat Prod Res 32(16):1926–1950. https://doi.org/10.1080/14786419.2017.1356838

Chen Y, Bilban M, Foster CA, Boger DL (2002) Solution-phase parallel synthesis of a pharmacophore library of HUN-7293 analogues: a general chemical mutagenesis approach to defining structure-function properties of naturally occurring cyclic (depsi)peptides. J Am Chem Soc 124(19):5431–5440. https://doi.org/10.1021/ja020166v

Chen Y, de Bruyn Kops C, Kirchmair J (2017) Data resources for the computer-guided discovery of bioactive natural products. J Chem Inf Model 57(9):2099–2111. https://doi.org/10.1021/acs.jcim.7b00341

Cozza G, Bonvini P, Zorzi E, Poletto G, Pagano MA, Sarno S et al (2006) Identification of ellagic acid as potent inhibitor of protein kinase CK2: a successful example of a virtual screening application. J Med Chem 49(8):2363–2366. https://doi.org/10.1021/jm060112m

Cragg GM, Newman DJ, Snader KM (1997) Natural products in drug discovery and development. J Nat Prod 60(1):52–60. https://doi.org/10.1021/np9604893

Dhiman P, Malik N, Khatkar A (2018) 3D-QSAR and in-silico studies of natural products and related derivatives as monoamine oxidase inhibitors. Curr Neuropharmacol 16(6):881–900. https://doi.org/10.2174/1570159X15666171128143650

Dias DA, Urban S, Roessner U (2012) A historical overview of natural products in drug discovery. Meta 2(2):303–336. https://doi.org/10.3390/metabo2020303

Grabowski K, Baringhaus K-H, Schneider G (2008) Scaffold diversity of natural products: inspiration for combinatorial library design. Nat Prod Rep 25(5):892–904. https://doi.org/10.1039/B715668P

Harvey AL, Edrada-Ebel R, Quinn RJ (2015) The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov 14(2):111–129. https://doi.org/10.1038/nrd4510

Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ et al (2015) Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 216:574–578

Khan MT, Orhan I, Senol FS, Kartal M, Sener B, Dvorska M et al (2009) Cholinesterase inhibitory activities of some flavonoid derivatives and chosen xanthone and their molecular docking studies. Chem Biol Interact 181(3):383–389. https://doi.org/10.1016/j.cbi.2009.06.024

Korotcov A, Tkachenko V, Russo DP, Ekins S (2017) Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. Mol Pharm 14(12):4462–4475. https://doi.org/10.1021/acs.molpharmaceut.7b00578

Lagunin AA, Goel RK, Gawande DY, Pahwa P, Gloriozova TA, Dmitriev AV et al (2014) Chemo- and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: a critical review. Nat Prod Rep 31(11):1585–1611. https://doi.org/10.1039/c4np00068d

Lavecchia A (2015) Machine-learning approaches in drug discovery: methods and applications. Drug Discov Today 20(3):318–331. https://doi.org/10.1016/j.drudis.2014.10.012

Leach AR, Gillet VJ, Lewis RA, Taylor R (2010) Three-dimensional pharmacophore methods in drug discovery. J Med Chem 53(2):539–558. https://doi.org/10.1021/jm900817u

Liu R, Li X, Lam KS (2017) Combinatorial chemistry in drug discovery. Curr Opin Chem Biol 38: 117–126. https://doi.org/10.1016/j.cbpa.2017.03.017

Liu M, Karuso P, Feng Y, Kellenberger E, Liu F, Wang C, Quinn RJ (2019) Is it time for artificial intelligence to predict the function of natural products based on 2D-structure. Medchemcomm 10(10):1667–1677. https://doi.org/10.1039/c9md00128j

Lusher SJ, McGuire R, van Schaik RC, Nicholson CD, de Vlieg J (2014) Data-driven medicinal chemistry in the era of big data. Drug Discov Today 19(7):859–868. https://doi.org/10.1016/j.drudis.2013.12.004

Ma D-L, Chan DS-H, Leung C-H (2011) Molecular docking for virtual screening of natural product databases. Chem Sci 2(9):1656–1665. https://doi.org/10.1039/C1SC00152C

Mang C, Jakupovic S, Schunk S, Ambrosi H-D, Schwarz O, Jakupovic J (2006) Natural products in combinatorial chemistry: an andrographolide-based library. J Comb Chem 8(2):268–274. https://doi.org/10.1021/cc050143n

McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB et al (2011) The eMERGE network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genet 4:13. https://doi.org/10.1186/1755-8794-4-13

Mladenovic M, Patsilinakos A, Pirolli A, Sabatino M, Ragno R (2017) Understanding the molecular determinant of reversible human monoamine oxidase B inhibitors containing 2H-Chromen-2-one core: structure-based and ligand-based derived three-dimensional quantitative structure-activity relationships predictive models. J Chem Inf Model 57(4):787–814. https://doi.org/10.1021/acs.jcim.6b00608

Naqvi AAT, Hassan MI (2017) Methods for docking and drug designing. In: Oncology: breakthroughs in research and practice. IGI Global, Philadelphia, pp 876–890

Naqvi AAT, Mohammad T, Hasan GM, Hassan MI (2018) Advancements in docking and molecular dynamics simulations towards ligand–receptor interactions and structure-function relationships. Curr Top Med Chem 18(20):1755–1768. https://doi.org/10.2174/1568026618666181025114157

Nasution MAF, Toepak EP, Alkaff AH, Tambunan USF (2018) Flexible docking-based molecular dynamics simulation of natural product compounds and Ebola virus Nucleocapsid (EBOV NP): a computational approach to discover new drug for combating Ebola. BMC Bioinform 19(suppl 14):419. https://doi.org/10.1186/s12859-018-2387-8

Newman DJ, Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79(3):629–661. https://doi.org/10.1021/acs.jnatprod.5b01055

Nguyen-Vo TH, Nguyen L, Do N, Nguyen TN, Trinh K, Cao H, Le L (2020) Plant metabolite databases: from herbal medicines to modern drug discovery. J Chem Inf Model 60(3): 1101–1110. https://doi.org/10.1021/acs.jcim.9b00826

Norn S, Permin H, Kruse PR, Kruse E (2009) From willow bark to acetylsalicylic acid. Dan Medicinhist Arbog 37:79–98

Onguene PA, Simoben CV, Fotso GW, Andrae-Marobela K, Khalid SA, Ngadjui BT et al (2018) In silico toxicity profiling of natural product compound libraries from African flora with anti-malarial and anti-HIV properties. Comput Biol Chem 72:136–149. https://doi.org/10.1016/j.compbiolchem.2017.12.002

Pathan H, Williams J (2012) Basic opioid pharmacology: an update. Br J Pain 6(1):11–16. https://doi.org/10.1177/2049463712438493

Patridge E, Gareiss P, Kinch MS, Hoyer D (2016) An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today 21(2):204–207. https://doi.org/10.1016/j.drudis.2015.01.009

Perez A, Martinez-Rosell G, De Fabritiis G (2018) Simulations meet machine learning in structural biology. Curr Opin Struct Biol 49:139–144. https://doi.org/10.1016/j.sbi.2018.02.004

Petrovska BB (2012) Historical review of medicinal plants' usage. Pharmacogn Rev 6(11):1–5. https://doi.org/10.4103/0973-7847.95849

Romano JD, Tatonetti NP (2019) Informatics and computational methods in natural product drug discovery: a review and perspectives. Front Genet 10:368. https://doi.org/10.3389/fgene.2019.00368

Shi C, Dong F, Zhao G, Zhu N, Lao X, Zheng H (2020) Applications of machine-learning methods for the discovery of NDM-1 inhibitors. Chem Biol Drug Des 96:1232. https://doi.org/10.1111/cbdd.13708

Tatonetti NP, Ye PP, Daneshjou R, Altman RB (2012) Data-driven prediction of drug effects and interactions. Sci Transl Med 4(125):125ra131. https://doi.org/10.1126/scitranslmed.3003377

Thomford NE, Senthebane DA, Rowe A, Munro D, Seele P, Maroyi A, Dzobo K (2018) Natural products for drug discovery in the 21st century: innovations for novel drug discovery. Int J Mol Sci 19(6):1578. https://doi.org/10.3390/ijms19061578

Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G et al (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 18(6):463–477. https://doi.org/10.1038/s41573-019-0024-5

Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design—a review. Curr Top Med Chem 10(1):95–115. https://doi.org/10.2174/156802610790232260

Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15(11–12):444–450. https://doi.org/10.1016/j.drudis.2010.03.013

Yao L, Zhang Y, Li Y, Sanseau P, Agarwal P (2011) Electronic health records: implications for drug discovery. Drug Discov Today 16(13–14):594–599. https://doi.org/10.1016/j.drudis.2011.05.009

Zhang DW, Luo RH, Xu L, Yang LM, Xu XS, Zheng YT, Luo H (2020) Natural-product-library-based screening for discovery of capsid C-terminal domain targeted HIV-1 inhibitors. Int J Antimicrob Agents 55(4):105926. https://doi.org/10.1016/j.ijantimicag.2020.105926

Zhu F, Li XX, Yang SY, Chen YZ (2018) Clinical success of drug targets prospectively predicted by in silico study. Trends Pharmacol Sci 39(3):229–231. https://doi.org/10.1016/j.tips.2017.12.002